JOURNAL OF UNDERGRADUATE RESEARCH

Predicting the Detectability of Safety Signals in Construction Sounds Using Machine Learning

Joseph Spagnoli[1], Aaron Costin[2]

[1]*Herbert Wertheim College of Engineering, University of Florida*
[2]*M.E. Rinker, Sr. School of Construction Management*

Mentor: Aaron Costin, M.E. Rinker, Sr. School of Construction Management

Abstract

Construction sites produce dynamic, multi-source noise that can mask safety-critical warning signals, contributing to struck-by fatalities. While psychoacoustic models such as Detectsound address masking prediction, they inadequately handle fluctuating noise characteristic of active construction, a gap identified by recent literature. This study investigates whether machine learning can predict warning signal audibility in construction noise. Eighty-one synthetic ISO 7731-compliant warning signals (tonal, multi-tone, and broadband) were mixed with real recorded construction noise at 10 signal-to-noise ratios (−10 to +30 dB), producing 56,700 audio mixtures labeled audible or masked via a two-tier scheme implementing ISO 7731 and ISO 9533 detectability criteria. Thirty-eight acoustic features including spectral, energy, MFCC, and frequency band measures were extracted per mixture. Four tree-based classifiers were evaluated. A Random Forest achieved 98.8% accuracy (F1 = 0.991), confirmed by 10-fold cross-validation (0.991 ± 0.002). The false positive rate, predicting "audible" when a signal is actually masked and the safety-critical error, was 1.19%. RMS energy variability and mid-frequency band power were the most discriminative features, consistent with ISO 9533 evaluation criteria. Convolutional neural network models trained on mel-spectrograms will be presented alongside these results as a deep learning comparison.

Keywords: auditory masking, construction safety, machine learning, warning signals, convolutional neural network

Introduction

Between 2011 and 2021, approximately 10.7 thousand construction workers died from occupational injuries, with struck-by incidents accounting for 17% of these fatalities (Harris et al., 2023). Auditory warning signals, such as vehicle backup alarms, serve as the primary mechanism to alert on-foot workers to the danger of approaching equipment (Alali & Casali, 2012). However, the effectiveness of these alarms is often reduced by high ambient noise levels, the varying hearing status of workers, and the use of hearing protection devices (Suter, 2002; Zheng et al., 2007).

Construction sites pose uniquely challenging acoustic environments characterized by varying and intermittent noises, unlike the more continuous noise typically found in manufacturing. Workers are routinely exposed to hazardous noise levels exceeding 85 dBA, with median exposures from equipment like pneumatic jackhammers exceeding 100 dBA (Suter, 2002). When this complex environment masks a warning signal, workers may fail to perceive approaching danger, leading to potentially fatal consequences (Zheng et al., 2007).

To address these challenges, Dang et al. (2022) conducted a systematic review examining the intersection of artificial intelligence, audio signal processing, and construction safety, finding that research on auditory situational awareness in construction is "still in its early stage, with inadequate AI-based sound sensing methods available" (p. 1). While previous machine learning research has primarily focused on classifying equipment activities, there remains a gap in predicting whether safety-critical warning signals remain audible to workers in noisy, dynamic construction environments.

Motivated by this gap, this study addresses the following research question: To what extent can machine learning techniques predict the detectability of safety warning signals under varying occupational soundscape conditions? Both classical machine learning (tree-based classifiers on engineered acoustic features) and deep learning (convolutional neural networks on mel-spectrograms) approaches are evaluated, with particular attention to the safety-critical false positive error of predicting a signal as audible when it is masked. The abstract for this work was based on a preliminary dataset of 56,700 mixtures; the final study expanded to 486,000 mixtures across three construction noise classes and completed CNN training, with all results reported herein reflecting the final analysis.

Methods

The methodology comprised four stages: (1) synthetic generation of 81 warning signals per ISO 7731 specifications, (2) selection and preparation of 600 construction noise segments from the UrbanSound8K dataset, (3) controlled audio mixing at 10 SNR levels with rule-based labeling implementing ISO 7731 and ISO 9533 detectability criteria, and (4) model development using engineered acoustic features for tree-based classifiers and mel-spectrograms for a convolutional neural network (**Figure 1**).

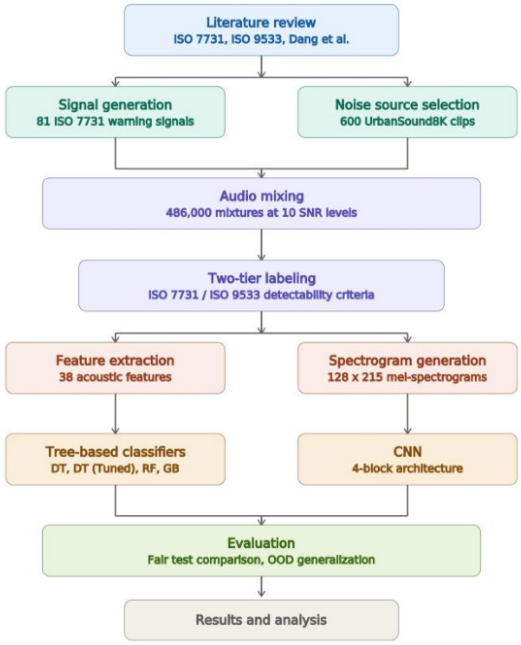

Figure 1. Research methodology pipeline flowchart. Synthesized warning signals and construction sounds were mixed, labeled per ISO criteria, processed for tree-based and CNN classification, and evaluated against each other.

Warning Signal Generation

No public dataset of construction-specific backup alarms exists. Existing urban sound datasets contain acoustically mismatched sources such as police sirens and car horns that differ fundamentally from the tonal, pulsed characteristics of construction backup alarms (Salamon et al., 2014). Warning signals were therefore synthesized programmatically in accordance with ISO 7731:2003 specifications for auditory danger signals. ISO 7731 specifies that danger signals should contain components within 500 to 2,500 Hz to ensure optimal detectability, should be intermittent to attract attention, and should differ clearly in temporal and spectral distribution from ambient noise. Three signal types were generated to cover the primary alarm categories recognized in the standard:

Tonal signals consisted of pulsed sine waves at five fundamental frequencies (600, 800, 1000, 1200, and 1500 Hz), spanning the ISO 7731-recommended range where human auditory sensitivity is highest (Laroche et al., 1991). Each frequency was combined with three on-times (0.3, 0.5, 0.75 s) and three off-times (0.2, 0.5, 1.25 s) matching ISO 7731 temporal specifications, yielding 45 unique tonal signals.

Multi-tone signals alternated between frequency pairs [(1000, 1200), (1100, 1300), (1200, 1400)] Hz on successive pulses, producing the distinctive alternating-frequency pattern characteristic of newer "smart" backup alarms. The same three on-time and off-time combinations yielded 27 signals. Multi-component signals are recommended by ISO 7731.

Broadband signals consisted of pulsed white noise bandpass-filtered to 500-2,500 Hz using a fourth-order Butterworth filter, with the same temporal patterns, yielding 9 signals. Broadband alarms improve localization over tonal alarms (Robin et al., 2020).

All 81 signals were generated at 22,050 Hz, 4.0 seconds in duration, with 5 ms linear fade-in/fade-out to prevent spectral artifacts from abrupt onset, and peak-normalized to 0.9 amplitude. The 4.0-second duration follows Salamon et al. (2014), who found 4 seconds sufficient for environmental sound identification with stable classification performance.

Noise Sources

Construction noise maskers were sourced exclusively from the UrbanSound8K dataset, a publicly available corpus of 8,732 labeled urban audio slices drawn from real field recordings on the Freesound platform (Salamon et al., 2014). Three classes acoustically representative of common construction noise conditions were selected: jackhammer (impulsive, broadband), drilling (sustained, mid-frequency), and engine idling (continuous, low-frequency). These classes represent the primary noise exposure categories in construction (impact tools, rotary tools, and heavy equipment; Suter, 2002) and span distinct spectral and temporal profiles relevant to masking prediction. Two hundred segments per class produced 600 total maskers. Each segment was loaded in mono at 22,050 Hz, trimmed or zero-padded to 4.0 seconds following the UrbanSound8K methodology for standardized slice duration, and saved as 16-bit PCM WAV.

Audio Mixing and Labeling

Each of the 81 warning signals was mixed with each of the 600 noise maskers at 10 signal-to-noise ratios (SNRs): -10, -6, -3, 0, +6, +10, +15, +20, +25, and +30 dB, producing 486,000 audio mixtures (81 × 600 × 10). SNR was computed on an RMS energy basis, with the signal scaled to achieve the target SNR relative to the masker's RMS level. Maskers were extended from 4.0 to 5.0 seconds by appending a segment from the middle of the original audio, and signals were placed at a randomized onset within the 5.0-second clip. The final mixture was peak-normalized to 0.9 amplitude to prevent clipping. The resulting dataset contained 322,085 audible samples (66.3%) and 163,915 masked samples (33.7%).

Each mixture was labeled using a two-tier scheme implementing detectability criteria from ISO 7731:2003 and ISO 9533:2010:

Tier 1 (A-weighted SNR): Mixtures with SNR above 15 dB were labeled audible; mixtures with SNR below 0 dB were labeled masked. Mixtures between 0 and 15 dB proceeded to Tier 2 analysis.

Tier 2 (1/3-octave band analysis): For borderline mixtures, signal and masker energies were computed in 1/3-octave bands at center frequencies from 500 to 4000 Hz. A signal was labeled audible if it exceeded the masker by more than 13 dB in any dominant frequency band, defined as any band within 6 dB of the signal's maximum band energy, per ISO 9533. Otherwise, the mixture was labeled masked.

Feature Extraction

Thirty-eight acoustic features were extracted from each mixture using librosa (version 0.11.0) with a frame size of 2,048 samples and hop length of 512 samples at 22,050 Hz. These feature categories are established descriptors in environmental sound classification (Dang et al., 2022; Salamon et al., 2014). The feature set comprised:

- *Spectral features* (15 features): mean and standard deviation of spectral centroid, spectral bandwidth, spectral rolloff, and spectral flatness (8 features), plus mean spectral contrast across seven sub-bands (7 features).
- *Energy features* (4 features): mean and standard deviation of RMS energy and zero-crossing rate. RMS standard deviation captures the temporal energy fluctuation characteristic of pulsed alarms, addressing the limitation of static level detection identified by Lutfi and Heo (2012).
- *Mel-frequency cepstral coefficients* (13 features): mean of MFCCs 1 through 13, computed with 128 mel bands. MFCCs are the most prevalent feature in sound classification research (Dang et al., 2022) and were used by Elelu et al. (2023) for CNN-based construction equipment detection.
- *Frequency band power* (6 features): power in six bands computed via FFT and converted to dB: low (0-500 Hz), low-mid (500-1,000 Hz), mid (1,000-2,000 Hz), high-mid (2,000-4,000 Hz), high (4,000-8,000 Hz), and very-high (8,000-11,025 Hz). The mid and high-mid bands align with the 500-4,000 Hz evaluation range specified by ISO 9533 Annex C.

No feature normalization was applied, as tree-based models are invariant to monotonic transformations (Breiman, 2001).

Model Architectures

tree-based classifiers.

Four tree-based classifiers were trained on the 38 features using scikit-learn (version 1.8.0). A baseline Decision Tree was trained without depth constraints. A tuned Decision Tree was optimized via GridSearchCV over max_depth [3, 5, 10, 15, 20, None], min_samples_split, and min_samples_leaf using 5-fold StratifiedGroupKFold. A Random Forest was trained with 100 trees, balanced class weights, and default hyperparameters. A Gradient Boosting classifier was trained with 100 estimators at the default learning rate of 0.1. All models were evaluated using 10-fold StratifiedGroupKFold cross-validation grouped by masker_id to prevent the same masker's spectral characteristics from appearing in both training and test folds.

convolutional neural network.

A CNN was trained on log-mel spectrograms of shape 128 × 215, computed with 128 mel bins, FFT size of 1,024, hop length of 512, and frequency range 200-8,000 Hz. Mel spectrograms provide a perceptually motivated time-frequency representation that enables CNNs to learn spectrotemporal features directly, an approach that has achieved strong results in environmental sound classification (Piczak, 2015; Zhang et al., 2018).

Spectrograms were globally z-score normalized (mean = -42.64, SD = 17.81) rather than per-sample, preserving absolute energy information critical to masking prediction. The architecture consisted of four convolutional blocks following the compact design pattern used by Elelu et al. (2023) for construction sound detection: Conv2d layers progressing from 1 to 32, 64, 128, and 128 channels (kernel size 3, padding 1), each followed by BatchNorm2d and ReLU, with MaxPool2d (kernel 2, stride 2) after the first three blocks. An AdaptiveAvgPool2d layer reduced the output to 128 × 1 × 1, followed by a fully connected layer (128 units, ReLU, dropout 0.5) and a single output logit. The model was trained using Adam (learning rate 1×10^{-3}) with ReduceLROnPlateau (patience 5, factor 0.5) and BCEWithLogitsLoss with pos_weight of 0.509 (n_masked / n_audible), upweighting the masked class in the loss function for safety conservatism. SpecAugment (Park et al., 2019) was applied during training with frequency masks of 5-15 mel bins and time masks of 10-30 frames. Training ran for up to 50 epochs with

early stopping (patience 10), converging at epoch 30. An 80/20 train-test split was implemented via GroupShuffleSplit (random_state = 42), grouped by masker_id.

Results

Tree-based Model Performance

Table 1 presents the performance of all four tree-based classifiers evaluated using 10-fold stratified grouped cross-validation on 486,000 mixtures.

Table 1. Tree-Based Classifier Performance (10-Fold Stratified Grouped Cross-Validation, N = 486,000)

Model	Accuracy	Precision	Recall	F1	CV Accuracy (SD)	CV F1 (SD)
Decision Tree (baseline)	.965	.966	.982	.974	.965 (.006)	.974 (.004)
Decision Tree (tuned)	.969	.976	.977	.976	.969 (.005)	.976 (.004)
Random Forest	.980	.977	.993	.985	.980 (.004)	.985 (.003)
Gradient Boosting	.969	.979	.974	.977	.969 (.005)	.977 (.004)

Note. All metrics are computed from out-of-fold predictions across 10 folds. SD = standard deviation across folds.

The Random Forest achieved the highest performance among tree-based models, with an F1 score of 0.985 and narrow cross-validation variance (SD = 0.003), indicating stable generalization across data folds. The baseline Decision Tree achieved 96.5% accuracy, demonstrating that even an unconstrained tree captures most of the audible-masked distinction. Gradient Boosting achieved accuracy comparable to the tuned Decision Tree (96.9%) but underperformed the Random Forest, likely due to insufficient hyperparameter optimization.

Fair Model Comparison

To compare the Random Forest and CNN on equal terms, both were evaluated on the identical 97,200-sample held-out test set produced by GroupShuffleSplit. Table 2 presents this comparison.

Table 2. Fair Comparison of Random Forest and CNN on Identical Test Set (n = 97,200)

Metric	Random Forest	CNN
Accuracy	.980	.985
Precision	.978	.994
Recall	.992	.983
F1 Score	.985	.989
AUC-ROC	.999	.999
FPR (safety-critical)	4.37%	1.19%
FNR	0.84%	1.68%

Note. FPR = false positive rate (predicted audible when masked). FNR = false negative rate (predicted masked when audible). Random Forest was retrained on the same 80/20 training split used by the CNN.

The CNN reduced the false positive rate by a factor of 3.7 relative to the Random Forest (1.19% vs. 4.37%). This came at the cost of a higher false negative rate (1.68% vs. 0.84%). In safety terms, the CNN's error profile is more conservative: it errs toward classifying signals as masked rather than audible, which represents the safer failure mode for worker protection. The CNN achieved slightly higher accuracy (98.5% vs. 98.0%) and substantially higher precision (99.4% vs. 97.8%), indicating fewer false alarms when predicting audibility.

Feature Importance Analysis

Feature importance rankings from the Random Forest revealed that RMS energy standard deviation (rms_std) dominated all other features with an importance score of 0.318, accounting for nearly one-third of the model's total predictive power. This feature captures temporal variability of signal energy across frames. When an alarm is audible, its periodic pulsing creates measurable fluctuations in the mixture's overall energy envelope. When the alarm is masked, these fluctuations are absorbed into the noise floor and rms_std decreases.

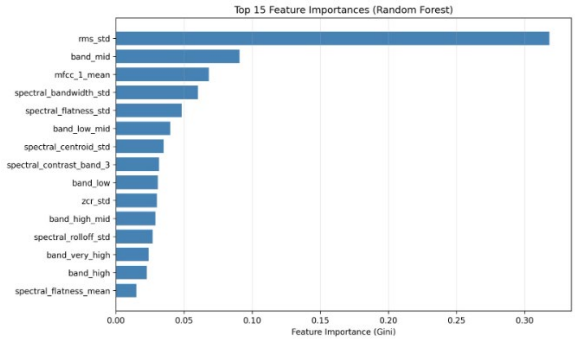

Figure 2. A bar graph illustrating top 15 Random Forest feature importances (Gini). RMS energy standard deviation (rms_std) was the dominant predictor, followed by mid-frequency band energy.

The second most important feature was mid-frequency band power (band_mid, importance = 0.091), which captures energy in the 1,000-2,000 Hz range where backup alarm fundamental frequencies typically reside. This aligns with ISO 9533:2010 evaluation criteria, which specify that audibility assessments should focus on frequency bands where warning signal energy is concentrated. Five of the top 10 features are temporal variability measures (standard deviations of RMS, spectral bandwidth, spectral flatness, and spectral centroid), reinforcing that the model distinguishes audible from masked signals primarily through temporal dynamics rather than static spectral shape.

Out-of-Distribution Generalization

To evaluate generalization beyond the training noise distribution, both models were tested on noise from the DEMAND database (Thiemann et al., 2013), which was never seen during training. Four subsets were tested: STRAFFIC (street traffic), TBUS (bus interior), TCAR (car interior), and TMETRO (metro interior). Ten randomly sampled signals were mixed with 20 noise clips per subset at five SNRs (-5 to +15 dB), producing 4,000 out-of-distribution mixtures.

Table 3. Out-of-Distribution Performance on DEMAND Noise (Never Seen During Training)

Subset	Model	Accuracy	F1	FPR	FNR
STRAFFIC	RF	.881	.930	.578	.001
STRAFFIC	CNN	.892	.937	.529	.000
TBUS	RF	.801	.889	.995	.000
TBUS	CNN	.800	.889	1.000	.000
TCAR	RF	.800	.889	1.000	.000

TCAR	CNN	.800	.889	1.000	.000
TMETRO	RF	.819	.898	.905	.000
TMETRO	CNN	.816	.897	.920	.000

Note. All subsets contain approximately 80% audible and 20% masked samples.

STRAFFIC (street traffic, approximately 70 dB) achieved the best out-of-distribution accuracy (approximately 89%) because its spectral profile most closely resembles the high-energy industrial noise in the training set. TBUS, TCAR, and TMETRO (transportation interiors, approximately 50-65 dB) exhibited false positive rates near 1.0, with both models classifying nearly all masked samples as audible. The training maskers (jackhammer at approximately 100 dB, drilling at approximately 95 dB, engine idling at approximately 80 dB) are substantially louder than transportation interior noise, so the models' learned association between low masker energy and signal audibility breaks down in quieter environments at low SNR. The near-zero false negative rates across all subsets demonstrate that neither model fails to identify an audible signal, maintaining conservative fail-safe behavior even under distribution shift.

Discussion

Model Performance and Error Analysis

Both models demonstrated that machine learning can predict warning signal detectability with accuracy exceeding 97%, supporting data-driven approaches as complements to existing psychoacoustic models. The progression from baseline Decision Tree (96.5%) through Random Forest (98.0%) illustrates ensemble variance reduction. The CNN's improvement (98.5% vs. 98.0%) is modest in absolute terms, but its error decomposition reveals a qualitatively different and safer failure mode.

The false positive rate is the safety-critical metric: predicting a signal is audible when actually masked. In deployment, this error means a worker is not alerted because the system incorrectly deemed the existing alarm sufficient. The CNN achieved 1.19% FPR compared to 4.37% for the Random Forest, a 3.7-fold reduction in the most dangerous error type. This improvement likely reflects the CNN's capacity to learn spatiotemporal patterns from mel-spectrograms that 38 scalar summary statistics cannot fully represent.

The CNN's higher false negative rate (1.68% vs. 0.84%) triggers unnecessary intervention but represents an acceptable safety tradeoff.

Alignment with Acoustic Standards

The feature importance analysis provides evidence that the models capture acoustically meaningful phenomena rather than statistical artifacts. The dominance of rms_std (31.8% importance) confirms that the Random Forest primarily detects the on-off cycling pattern of pulsed alarm signals, consistent with the temporal-envelope mechanism described in the Results.

The importance of mid-frequency band power (band_mid) aligns with ISO 9533:2010 performance criteria, which evaluate earth-moving machinery alarms in frequency bands where backup alarm fundamental frequencies and their harmonics are concentrated. The contribution of spectral contrast features, particularly in band 3 (corresponding to mid frequencies), mirrors the psychoacoustic principle that signal detection depends on the spectral contrast between signal and noise within overlapping critical bands (Laroche et al., 1991).

Out-of-Distribution Generalization

The out-of-distribution results highlight an important boundary condition. Both models maintained near-perfect recall on unseen noise types but showed elevated false positive rates for quieter transportation interiors (TBUS, TCAR, TMETRO). This reflects domain mismatch rather than model failure: training on high-energy industrial noise taught that low masker energy predicts audibility, an inference that breaks down in quieter environments at low SNR. Masking prediction models are therefore environment-specific, requiring representative training data for deployment.

Limitations

Several limitations of this work should be acknowledged. Warning signals were generated synthetically per ISO 7731 specifications rather than recorded from actual construction equipment. While this ensures controlled signal properties, real backup alarms exhibit manufacturing variability, speaker degradation, and environmental propagation effects not captured in synthetic signals.

The labeling scheme applies ISO 7731 and ISO 9533 detectability criteria as a computational oracle, but these standards were developed primarily for steady-state conditions and have not been validated against human perception in fluctuating noise. The high model accuracy therefore reflects the models' ability to learn the rule-based labeling function, not necessarily their ability to predict what a human worker would hear.

All audio mixtures were generated under controlled conditions without accounting for spatial propagation, reverberation, hearing protection attenuation, or individual hearing loss. Construction workers frequently wear hearing protection and may have noise-induced hearing loss, both of which significantly affect detection thresholds (Laroche et al., 1991; van der Hoek-Snieders et al., 2020).

Finally, the out-of-distribution results demonstrate that models trained on industrial construction noise do not generalize to substantially different noise environments without retraining. Deployment in specific construction contexts would require noise data representative of those contexts.

Conclusion

This study demonstrates that machine learning can predict safety warning signal detectability in construction noise with accuracy exceeding 98%. A Random Forest classifier trained on 38 engineered acoustic features achieved 97.95% accuracy (F1 = 0.985), while a convolutional neural network trained on mel-spectrograms achieved 98.49% accuracy (F1 = 0.989) with a safety-critical false positive rate of 1.19% on the same held-out test set. The feature importance analysis revealed that temporal energy variability and mid-frequency band power are the most discriminative predictors, consistent with ISO 9533 evaluation criteria and psychoacoustic masking principles.

These results address a gap identified by Dang et al. (2022) by demonstrating the feasibility of data-driven masking prediction for construction warning signals. The approach complements existing psychoacoustic models by accommodating the temporal variability of construction noise that steady-state models cannot capture. Out-of-distribution testing confirmed that both models maintain conservative fail-safe behavior on unseen noise types, though false positive rates increased substantially outside the training noise distribution, indicating that deployment would require environment-specific training data.

Future work should pursue three directions. First, model predictions should be validated against human listener studies conducted under controlled noise conditions and, ultimately, in field settings on active construction sites. Second, the training noise distribution should be expanded to include a broader range of construction-specific noise sources, including pile driving, concrete pumping, and demolition, to improve out-of-distribution generalization. Third, integrating worker-specific factors such as hearing protection attenuation and age-related hearing

loss would move the approach toward a practical system for real-time audibility assessment on construction sites.

Acknowledgments

This work was supported by the University of Florida AI Scholars Program. Special thanks to Dr. Aaron Costin for his mentorship throughout this project and to Dr. Hassan Azad for identifying candidate noise datasets that informed the data selection for this research. A final thanks to Lauren Nunag for her advice and guidance on navigating undergraduate research.

References

Alali, K., & Casali, J. G. (2012). Auditory backup alarms: distance-at-first-detection via in-situ experimentation on alarm design and hearing protection effects. *WORK: A Journal of Prevention, Assessment & Rehabilitation*, *41*(S1), 3599–3607. https://doi.org/10.3233/wor-2012-0671-3599

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Dang, K., Elelu, K., Le, T., & Le, C. (2022). Augmented Hearing of Auditory Safety Cues for Construction Workers: A Systematic Literature Review. *Sensors*, *22*(23), 9135. https://doi.org/10.3390/s22239135

Elelu, K., Le, T., & Le, C. (2023). Collision Hazard Detection for Construction Worker Safety Using Audio Surveillance. *Journal of Construction Engineering and Management*, *149*(1). https://doi.org/10.1061/jcemd4.coeng-12561

Harris, W., Yohannes, T., & Trueblood, B. (2023). Fatal and Nonfatal Focus Four Injuries in Construction. In *CPWR*. https://www.cpwr.com/wp-content/uploads/DataBulletin-March2023.pdf

International Organization for Standardization. (2008). *Ergonomics — Danger signals for public and work areas — Auditory danger signals* (ISO Standard No. 7731:2003). https://www.iso.org/standard/33590.html

International Organization for Standardization. (2010). *Earth-moving machinery — Machine-mounted audible travel alarms and forward horns — Test methods and performance criteria* (ISO Standard No. 9533:2010).

Laroche, C., Tran Quoc, H., Hétu, R., & McDuff, S. (1991). "Detectsound": A computerized model for predicting the detectability of warning signals in noisy workplaces. *Applied Acoustics*, *32*(3), 193–214. https://doi.org/10.1016/0003-682x(91)90003-w

Lutfi, R., & Heo, I. (2012). Automated detection of alarm sounds. *The Journal of the Acoustical Society of America*, *132*(2), EL125-8. https://doi.org/10.1121/1.4734555

Park, D., Chan, W., Zhang, Y., Chiu, C.-C., Zoph, B., Cubuk, E., & Le, Q. (2019). *SpecAugment: A simple data augmentation method for automatic speech recognition*. *arXiv*. https://arxiv.org/abs/1904.08779

Piczak, K. J. (2015). Environmental sound classification with convolutional neural networks. In *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1–6). IEEE. https://doi.org/10.1109/mlsp.2015.7324337

Robin, O., Krpic, T., Nélisse, H., & Berry, A. (2020). Acoustic characterization of tonal and broadband backup alarms in laboratory and field conditions. *Applied Acoustics*, *163*, 107228. https://doi.org/10.1016/j.apacoust.2020.107228

Salamon, J., Jacoby, C., & Bello, J. P. (2014). *A dataset and taxonomy for urban sound research*. In Proceedings of the 22nd ACM International Conference on Multimedia (pp. 1041–1044). ACM. https://doi.org/10.1145/2647868.2655045

Suter, A. H. (2002). Construction Noise: Exposure, Effects, and the Potential for Remediation; A Review and Analysis. *AIHA Journal*, *63*(6), 768–789. https://doi.org/10.1080/15428110208984768

Thiemann, J., Ito, N., & Vincent, E. (2013). The Diverse Environments Multichannel Acoustic Noise Database (DEMAND): A database of multichannel environmental noise recordings. *Proc. Mtgs. Acoust.*, *19*(1), 035081. https://doi.org/10.1121/1.4799597

van der Hoek-Snieders, H. E. M., Houben, R., & Dreschler, W. A. (2020). Detectability of auditory warning signals in the ambient noise of Dutch train cabins. *Ergonomics*, *64*(4), 474–484. https://doi.org/10.1080/00140139.2020.1837397

Zhang, Z., Xu, S., Cao, S., & Zhang, S. (2018). *Deep convolutional neural network with mixup for environmental sound classification*. *arXiv*. https://arxiv.org/abs/1808.08405

Zheng, Y., Giguère, C., Laroche, C., Sabourin, C., Gagné, A., & Elyea, M. (2007). A Psychoacoustical Model for Specifying the Level and Spectrum of Acoustic Warning Signals in the Workplace. *Journal of Occupational and Environmental Hygiene*, *4*(2), 87–98. https://doi.org/10.1080/15459620601115768